

HOW RONALD FISHER BECAME A MATHEMATICAL STATISTICIAN

Stephen M. STIGLER¹

RÉSUMÉ – Comment Ronald Fisher devint statisticien

En hommage à Bernard Bru, cet article rend compte de l'influence décisive qu'a eu Karl Pearson sur Ronald Fisher. Fisher ayant soumis un bref article rédigé de façon hâtive et insuffisamment réfléchi, Pearson lui fit réponse éditoriale d'une grande finesse qui arriva à point nommé.

MOTS-CLÉS – Exhaustivité, Inférence fiduciaire, Maximum de vraisemblance, Paramètre

SUMMARY – *In homage to Bernard Bru, the story is told of the crucial influence Karl Pearson had on Ronald Fisher through a timely and perceptive editorial reply to a hasty and insufficiently considered short submission by Fisher.*

KEY-WORDS – Fiducial inference, Maximum likelihood, Parameter, Sufficiency,

It is a great personal pleasure to join in this celebration of the friendship and scholarship of Bernard Bru. Over a period of more than 20 years I have enjoyed adventures with Bernard, adventures that have taken us to surprising and exotic corners of the literature of the statistical world. We have chased ghosts in that literature – references to works that cannot be proven to exist, nor can they be proven not to exist. We have sought skeletons, as in the literature of craniometry. We have pursued foods and recipes, as in the case of the dessert invented by Laplace and made famous by Brillat-Savarin. He has helped me identify rare works of art – such as a contemporary drawing of Laplace and Napoleon. He has identified for me inscriptions on the Pantheon that are barely a trace, and obscure citations that would have stymied Inspectors Poirot and Maigret. In his *Essai philosophique sur les probabilités*, Laplace famously described a deterministic world where to an infinite intelligence nothing would be beyond the view, a description that reasonably well captures Bernard's vast knowledge of the great historic literature of probability and statistics.

On this occasion I wish to discuss an unusual event in the early history of twentieth century statistics, and to share with you what was for me a surprising discovery. It involves Ronald Fisher and Karl Pearson, and it may help explain a key point in the development of modern statistics. And it has a tangential but significant link to research of Bernard's that goes forward at present.

¹ Department of Statistics and the Committee on Conceptual and Historical Studies of Sciences, University of Chicago, stigler@galton.uchicago.edu

Fisher has always been a puzzle to modern statisticians. The vast influence of his work on experimental design is undeniable. The impact of his methods and theory, from his analysis of variance to his distributions of correlation coefficients and regression estimates, is found everywhere in our texts and courses. But other aspects of Fisher's work and thought are more elusive. What is fiducial inference? What exactly is an ancillary statistic? What is a recognizable subset? Under what conditions are maximum likelihood estimates truly optimal? In the first half of the twentieth century, the great Maurice Frechet was among those who recognized the fertility of Fisher's work but also found some parts of it mysterious. While most people were content to scratch their heads and make quiet comments about their doubts that Fisher had a clear idea in mind on all of these, Frechet took action. In the 1930s, Frechet had generated a considerable correspondence toward the goal of better understanding a Karl Pearsonian structure, the correlation coefficient. Much of this was published at the time in the journals of the International Statistical Institute. With the fruit born of that dialogue, it is understandable that in the 1940s Frechet undertook a different and altogether more challenging attempt to better understand fiducial inference. And as Bernard Bru's deep and ongoing work on French work of this period went forward, it is natural that he would uncover the results of Frechet's efforts in the Frechet archive in the Académie des Sciences.

Some of this correspondence has since been published with a volume of Fisher's correspondence [Bennett, 1990], but the correspondents ranged much further than just Fisher, as Frechet sought to draw out a wide spectrum of people on an elusive idea, although in the end he failed to make significant headway. The question Frechet asked is not one I will address here, except indirectly perhaps. But Bernard shared this archive with me at a time when I was beginning to look seriously at Fisher's work as a puzzled historian instead of as a puzzled practicing statistician, and that archive helped condition me to what I think is a new understanding of an earlier event that was crucial to Fisher's development as a statistician. It is that event that I will discuss here, the question of how Fisher became a statistician in the first place. I tell the story in more detail elsewhere [Stigler, 2005]; my account here will concentrate upon the narrative, and, yes, the drama of the event.

It was by no means foreordained that Fisher would work in statistics at all. He had a brilliant career as a mathematics student at Cambridge University and published two student papers at that time, one a short piece on statistical error theory, the other a longer exposition on vector analysis through differential geometry. Neither advanced their subjects in important ways, but the second showed a firm grasp of an astonishing range of modern mathematics. After he left Cambridge in 1913, he largely dropped out of mathematics, as he became an indifferent teacher in a series of secondary schools and devoted his intellectual energy almost exclusively to eugenics, both social and scientific issues related to issues of heredity, with particular attention to how the latter could inform the former. Through the years 1914 to 1920 he published 95 separate pieces. Of these, 92 were on eugenics (mostly book reviews), one was related to eugenics concerns (his major 1918 work on statistical genetics), and only two could be said to be centrally concerned with mathematical statistics. One of these was his masterful derivation of the distribution of the correlation coefficient in 1915; the other a work that appeared in 1920 and is related to what I will discuss here. This was not on the face of it the record of a major statistician, nor for that matter a dedicated geneticist. And yet in 1921 Fisher presented a major pathbreaking memoir to the Royal Society of London, a memoir that more than any other single work gave the framework and direction to twentieth century statistical theory.

Fisher's memoir [Fisher, 1922] introduced ideas as basic as the parametric family of distributions (the word "parameter" appears 57 times in the memoir, its first appearance in the modern statistical sense). The terms and concepts of likelihood, sufficiency, consistency, efficiency are new here, too. The memoir focused upon the theory of estimation, but a few years later Jerzy Neyman and Egon Pearson adopted Fisher's framework for their theory of hypothesis testing, and the resurgence of Bayesian statistics later in that century also adopted Fisher's framework, even if not his use of it. The memoir was in this way a watershed for twentieth century statistics. It had nothing to do with eugenics, Fisher's principal preoccupation at that time. Where did it come from?

Recall that through this decade Fisher, trained as a mathematician, was most deeply concerned with eugenics. A good share of this interest was social policy-oriented, but the scientific issues caught his attention as well. And at that time anyone with a mathematical interest in eugenics would have naturally been drawn to the pages of the journal *Biometrika*, founded in 1901 by Karl Pearson. Indeed, one of Fisher's notes in a eugenics journal in 1915 was simply to point out to the eugenics community the importance of *Biometrika* to eugenics. Once his eyes were on *Biometrika*, they inevitably strayed to some more basic articles. As early as 1912 he had noted "Student" writing in *Biometrika* on the t-test, and Fisher had sent him a cryptic letter suggesting how the t-distribution could be rigorously derived. And in 1914 Fisher similarly was moved, this time by an article by H. E. Soper trying to approximate the distribution of the correlation coefficient. This time his interest resulted in a publication, his first important contribution to statistics [Fisher, 1915]. Still, as impressive as his derivation of the distribution of the sample correlation coefficient was, with its mastery of high dimensional spaces that had been foreshadowed by the 1913 college paper on vector analysis, it still did not hint at the author of Fisher [1922]. The article showed Fisher's energy and mathematical facility, and that he could solve technical problems beyond any other statisticians of that time, but there was no sign of the theory, the statistical conceptualization that was to come.

The crucial event I describe came in 1916. Fisher was continuing to think a bit about the correlation coefficient, and his relations with Karl Pearson remained cordial. Fisher had also started work on his first major effort on statistical genetics. This work, on the correlation of relatives under Mendelian inheritance, was eventually published in 1918, and may be seen as broadly within his eugenics program. The techniques he used carried the seeds of later work in the analysis of variance (and much other work in genetics) but it too lacked a framework within statistical theory. In May of 1916, that year's only issue of *Biometrika* appeared, and within a month Fisher wrote a short note to Pearson, with two enclosures. One of these was a draft of the work on Mendelian inheritance, a dense and hard to read memoir, even today. Fisher knew Pearson was interested in this topic, even if he probably did not expect Pearson to take to his findings immediately. He would have known that Pearson and the Mendelian biologist Bateson had quarreled earlier on the value of biometry, and Pearson was on record as doubting that Mendelian inheritance could easily explain inheritance in human populations [Porter, 2004]. Fisher's work would reconcile the positions, once it was understood and accepted. The other enclosure was in comparison a trivial effort. It was only one page in length, but unlike the long genetics article, it was sent as a submission for publication in *Biometrika*. Fisher wrote,

There is an article by Miss Kirstine Smith in the current number of Biometrika which, I think, ought not to pass without comment. I enclose a short note on it.

Fisher's letter, his one page submission, and Pearson's reply are all reprinted in full in a 1968 article by E. S. Pearson, and various portions of them have been cited and quoted by several subsequent authors. Yet I think the full significance of this exchange has been missed. On the face of it, the issue was a simple one. The paper referred to [Smith, 1916] was by a young Dane working as a doctoral student in Pearson's laboratory. She presented through examples a short case for the use of minimum chi-square as a criterion for estimation, in preference to what she called the Gaussian method. The Gaussian method was the then-classical staple of error theory; it was essentially maximum likelihood, based upon a vague Bayesian argument where a uniform prior was assumed for the object of estimation. Smith thought the use of a distance measure (the chi-squared metric) was preferable, even though it required grouped data. Some of her examples used discrete data, some involved grouped continuous data. Fisher's one-page hand-written submission complained that with continuous data, different ways of grouping the data would produce different results, a logical inconsistency that should be avoided. As Fisher wrote :

There is nothing at all 'arbitrary' in the use of the method of moments for the normal curve; as I have shown elsewhere it flows directly from the absolute criterion ($\Sigma \log f$ a maximum) derived from the Principle of Inverse Probability. There is, on the other hand, something exceedingly arbitrary in a criterion which depends entirely upon the manner in which the data happen to be grouped.

Fisher's reference to having "shown elsewhere" was to his undergraduate paper, Fisher [1912], where he had used the Gaussian method in choosing a normal distribution from data, where it led to method-of-moments values [Aldrich, 1997; Edwards, 1997; Hald, 1998]. Pearson also had used the Gaussian method in 1898 to derive the estimate of the correlation coefficient, and Fisher was apparently defending a Pearsonian position in his submission. The response he got must have been an unwelcome surprise.

Pearson's reply was a full two pages – longer than the submission it declined. The temptation is to read it as a dismissal of a paper of merit that criticized work from Pearson's own laboratory. Pearson did do that, but not on the simple basis that is tacit in that description. This was not the letter of a closed-minded older man unwilling to abide with even indirect criticism. It was rather the letter of a first-rate editor to a novice author who had sent a paper in without seriously considering the issues involved. Pearson told Fisher exactly what was wrong with his note, and exactly what was needed to correct it.

Fisher had adopted the Gaussian method without serious examination of the principles behind it. He seems to have been attracted to it only as a handy tool that gave results that seemed to be free of mathematical inconsistencies such as he noted in Smith's paper. He seemed to have never considered the case that needed to be presented for his own position.

Pearson noted in his letter that he himself had used the Gaussian method before [Pearson, Filon, 1898], but he now doubted it on logical grounds. For the main philosophical objection to the Gaussian method, he referred to a footnote in Smith's paper where she noted the probability being maximized by the Gaussian method was

hardly a probability at all; and even if it be multiplied by a differential element to make it a probability of a neighborhood of the observed data, that probability was infinitely small. What use was a criterion that gave a result with an infinitely small probability? Much better to use a distance measure like chi-square. Pearson closed with an invitation.

If you will write me a defence of the Gaussian method, I will certainly consider its publication, but if I were to publish your note, it would have to be followed by another note saying that it missed the point, and that would be a quarrel among contributors.

There is no reply from Fisher in the archives. But what could he have said? The truth is that there was at that time no principled reply. Fisher, like a host of others before him, had accepted a vague inverse-probability appeal as sufficient to justify the maximization of a probability or density that even at its maximum is very small. By the cold light of Pearson's straightforward letter, there really was no case. Fisher put the letter aside (it was among a small amount of early correspondence he preserved to the end of his life), grumbled silently about Pearson, and returned to other business.

But he must have continued to think about the issue Pearson raised. In 1919, shortly before he moved from teaching to a research post at Rothamsted's agricultural research station, he was led by chance to an astounding discovery. The astrophysicist A. S. Eddington had claimed in a 1914 book that it was better to estimate the standard deviation of a normal distribution using a multiple of the mean absolute deviation from the mean, than using a multiple of the root mean squared deviation from the mean, adding,

This is contrary to the advice of most text-books; but it can be shown to be true [Eddington, 1914, p. 147].

Fisher's skills with multiple algebra made short work of Eddington's erroneous claim: in small and in large samples, the usual estimate had a smaller standard deviation. He wrote to Eddington, and Eddington replied in July 1919, agreeing that his statement had been wrong. But Fisher did not stop there; he took one more step, and that step was to prove decisive. For the case $n = 4$, Fisher looked to see if perhaps even though the usual estimate was superior, some improvement might be had by incorporating the other estimate in some way, say using an average of the two. He derived the conditional distribution of Eddington's estimate given the usual one and was astounded to find it did not involve the object of estimation at all! Once he had the usual estimate in hand there was absolutely no prospect of gain from the other estimate! When Fisher published this little investigation in 1920, his italicized statement fairly shouted this result:

The whole of the information respecting [the standard deviation], which a sample provides, is summed up in the value of [the usual estimate].

This was his first discovery of sufficiency [Fisher, 1920; Stigler, 1973].

Here at last he had a germ of a reply to Pearson. The next step he took is not recorded, but it must have been to ask how general this phenomenon was. I speculate that what he did next is precisely the first mathematical derivation he presented in the 1922 memoir. There he assumed that he had two estimates, one a sufficient statistic, and the two of them having approximately a bivariate normal distribution. Working with that bivariate density he then showed that the insufficient estimate must have a larger standard deviation than the sufficient estimate. At that time he was convinced (wrongly,

as it turned out) that the Gaussian method, which he now called the method of maximum likelihood, always produced a sufficient estimate. If that was so, he had a devastating reply to Pearson. Maximizing the likelihood function might not produce a very high value for that probability or density, but for large samples it would automatically lead to the estimate with the smallest standard deviation.

He submitted his memoir to the Royal Society in June 1921. So much intricate work was involved that he would have had little time for anything else, although he did also review 17 books on eugenics that year. The memoir examined in detail the relative worth of Pearson's method of moments and the newly named method of maximum likelihood, showing that within the Pearson family of densities, the method of moments only performed decently for curves near the normal density, where it was in fact maximum likelihood. There was much more in the memoir, including a look at the method of minimum chi-square (it was equivalent to maximum likelihood for large samples, but suffered a second order loss of information for finite samples).

Pearson's 1916 challenge had born remarkable fruit – the modern theory of estimation and the associated conceptual structure were effectively born in that single memoir. Fisher however never acknowledged Pearson as an important influence on him, and Pearson never saw that Fisher's theory was an effective answer to the challenge of 1916. And one of the first bonuses from Fisher's parametric framework led to Fisher's correction of Pearson's error on the number of degrees of freedom for the chi-square test, a discovery that did not help the relationship between the two [Stigler, 1999(a), Chapter 19]. Indeed, the impact of Fisher's memoir was only slowly felt. Fisher himself rewrote and expanded portions of it [Hinkley, 1980], and the framework was further spread through its adoption by Neyman and Egon Pearson in 1928. He achieved early recognition in America by Harold Hotelling [Stigler, 1999(b)]. The proofs Fisher gave were incomplete and vague in detail, and debate on central points extended for decades. But the statistical landscape was forever changed.

The point I emphasize here is that it was only when Pearson challenged him in the highest tradition of scientific editing in 1916 that Fisher paused to consider the case to be made for his methods, and saw that it was deficient. His biggest influence at this stage was Pearson; not eugenics, not the research environment at Rothamsted, not his teaching or university training, but Karl Pearson. Without Pearson's question, it seems extremely unlikely Fisher would have seen how his surprising discovery of sufficiency was the key to a whole theory of estimation, and a broad framework for addressing statistical questions more broadly.

Does this event have any bearing on the question Frechet asked in the 1940s about the meaning of fiducial inference? Perhaps, even if only through its indication of the way Fisher's mind worked, in making extraordinarily fruitful leaps that could not always be sustained under rigorous mathematical examination. Fisher's early enthusiastic assumption that maximum likelihood estimates were always sufficient was put aside later, leaving holes that were never fully addressed. His claims about ancillary statistics sometimes exceeded what he or anyone else could show [Stigler, 2001]. The history of our subject is filled with fascinating mysteries. Bernard Bru understands this better than anyone, and his work has deeply enriched us. I regard these remarks as influenced by the curiosity aroused through his sharing of the Frechet archival discovery, and I am, unlike Fisher, pleased to acknowledge his assistance. I eagerly await his completion of his great work-in-progress on twentieth century French probability and statistics.

REFERENCES

- ALDRICH J., "R. A. Fisher and the making of maximum likelihood 1912-1922", *Statistical Science* 12, 1997, p. 162-176.
- BENNETT J. H. (ed), *Statistical inference and analysis: selected correspondence of R. A. Fisher*, Oxford, Clarendon Press, 1990.
- BOX, J. F., *R. A. Fisher: the life of a scientist*, New York, Wiley, 1978.
- EDDINGTON A. S., *Stellar movements and the structure of the universe*, London, Macmillan, 1914.
- EDWARDS, A.W.F., "What did Fisher mean by 'inverse probability' in 1912-1922?", *Statistical Science* 12, 1997, p. 177-184.
- FIENBERG S.E., HINKLEY D.V., (eds), "R. A. Fisher: an appreciation", published as *Lecture Notes in Statistics*, Volume 1, New York, Springer-Verlag, 1980.
- FISHER R.A., "On an absolute criterion for fitting frequency curves", *Messenger of Mathematics* 41, 1912, p. 155-160; [reprinted as Paper 1 in Fisher, 1974; reprinted in Edwards, 1997].
- FISHER R.A., "Applications of vector analysis to geometry", *Messenger of Mathematics* 42, 1913, p. 161-178; [reprinted as Paper 2 in Fisher, 1974].
- FISHER R.A., "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population", *Biometrika* 10, 1915, p. 507-521; [reprinted as Paper 4 in Fisher, 1974].
- FISHER R.A., "The correlation between relatives on the supposition of Mendelian inheritance", *Transactions of the Royal Society of Edinburgh* 52, 1918, p. 399-433; [reprinted as Paper 9 in Fisher, 1974. Also reprinted with commentary in Moran and Smith, 1966].
- FISHER R.A., "A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error", *Monthly Notices of the Royal Astronomical Society* 80, 1920, p. 758-770; [reprinted as Paper 12 in Fisher, 1974].
- FISHER R.A., "On the mathematical foundations of theoretical statistics" *Philosophical Transactions of the Royal Society of London (A)* 222, 1922, p. 309-368. [Reprinted as Paper 18 in Fisher, 1974].
- FISHER R.A., *The collected papers of R. A. Fisher* (ed. J. H. Bennett), Adelaide, University of Adelaide Press, 1974.
- GEISSER S., "Introduction to Fisher (1922). On the mathematical foundations of theoretical statistics", *Breakthroughs in Statistics*, Vol. I, (S. Kotz and N. L. Johnson, eds), New York, Springer-Verlag, 1992.
- HALD A., *A history of mathematical statistics from 1750 to 1930*, New York, Wiley, 1998.
- HINKLEY D.V., *Theory of statistical estimation: The 1925 paper*, 1980, p. 85-94.
- KENDALL M.G., PLACKETT R.L., *Studies in the history of statistics and probability*, Vol. 2, London, Griffin, 1977.
- MORAN P.A.P., SMITH C.A.B., *Commentary on R. A. Fisher's paper on 'The correlation between relatives on the supposition of Mendelian inheritance'*, Eugenics Laboratory Memoirs XLI, published for the Galton Laboratory, University College London, by Cambridge University Press, 1966.
- NEYMAN J., PEARSON E.S., "On the use and interpretation of certain test criteria for purposes of statistical inference" Part I, *Biometrika* 20A, 1928, p. 175-240.

- PEARSON E.S., "Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with notes and comments", *Biometrika* 55, 1968, p. 445-457.
- PEARSON K., FILON L.N.G., "Mathematical contributions to the theory of evolution IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation", *Philosophical Transactions of the Royal Society of London*, Series A, 191, 1898, p. 229-251.
- PORTER Th.M., *Karl Pearson: The Scientific Life in a Statistical Age*, Princeton, Princeton University Press, 2004.
- SMITH K., "On the 'best' values of the constants in frequency distributions", *Biometrika* 11, 1916, p. 262-276.
- STIGLER S.M., "Laplace, Fisher, and the discovery of the concept of sufficiency", *Biometrika* 60, 1973, p. 439-445, [reprinted in Kendall and Plackett, 1977, p. 271-277].
- STIGLER S.M., Contribution to discussion of "On rereading R. A. Fisher," by L. J. Savage, *Annals of Statistics* 4, 1976, p. 498-500.
- STIGLER S.M., *The history of statistics: the measurement of uncertainty before 1900*, Cambridge (Mass), Harvard University Press, 1986.
- STIGLER S.M., *Statistics on the table*, Cambridge (Mass), Harvard University Press, 1999(a).
- STIGLER S.M., "The foundations of statistics at Stanford" *American Statistician* 53, 1999(b), p. 263-266.
- STIGLER S.M., "Ancillary history", M.C.M. de Gunst, C.A.J. Klaassen, A.W. van der Vaart (eds), *State of the art in probability and statistics; festschrift for Willem R. van Zwet*, Institute of Mathematical Statistics, Lecture Notes-Monograph Series 36, 2001, p. 555-567.
- STIGLER S.M., "Fisher in 1921", *Statistical Science* 20, 2005, p. 32-49.
- ZABELL S.L., "R. A. Fisher and the fiducial argument", *Statistical Science* 7, 1992, p. 369-387.